Verification and Validation of Models and Simulations

(Excerpted from an 2001 NASA proposal written by Ken Leiden, Ron Laughery, and Kevin Corker)

2.3.2.3 Validation Requirements

V&V involves the *verification* phase, where it is determined that the software behaves as it is claimed (e.g., algorithms are implemented correctly, random number generators produce truly random numbers). The *validation* phase focuses on the ability of the model to provide sound predictions. Central to validation is defining the scope of the issues that the model/simulation components can and can not address. Since validation can be an extremely tedious process, the implications of scope and fidelity of AMSAT on the required validation should not be underestimated. The selection of performance measures that AMSAT must predict should be limited to the specific questions that are being asked.

2.3.2.3.1 Verification

Because AMSAT is a suite of simulations/models, the verification phase has some extra steps as compared to verification for stand-alone simulations. The following list documents the six-step process towards AMSAT verification:

1. All models/simulations to be considered usable components of AMSAT are verified in a stand-alone sense to the extent that this is reasonable and feasible. This level of verification will be performed by the CSC team member that is most familiar with the model/simulation. This would typically involve creating realistic test cases that exercise all the lines of code in the model/simulation component and ensuring that the results are correct.
2. The interface between each simulation pair is tested for correctness for all possible message types as defined in the interface requirements specification (Section 2.3.3.2). This is done for every simulation pair that is required to communicate for the AMSAT study being addressed. This approach applies to both sequential models (one models output is the other model's input) and dynamic models.
3. Each model/simulation is tested to determine that the parameters being sent or received from the interface are being utilized correctly by the internal algorithms of the model/simulation.
4. For dynamic simulations, the synchronization and management of *simulation time* must be demonstrated. This is particularly important for dynamic simulation pairs in which one is a discrete event simulation and the other is a time-step driven simulation.
5. The suite of models/simulations for a particular AMSAT scenario are run for the duration of that scenario. Extensive data collection is gathered during this run by all models/simulations. The CSC team member that is most familiar with a given model/simulation will be responsible for extensive post-run analysis.
6. Lastly, if errors are found in step 5, the AMSAT scenario is re-run until all analysts/modelers confirm that the scenario was executed to completion without errors.

One of the goals of this CTO is to develop a simulation toolkit where the level of model fidelity is chosen based on the question being asked. When a model component is being replaced by a component of a different level of fidelity (e.g., low fidelity point mass aircraft model vs. high fidelity 6 DOF aircraft model), the verification process must be repeated to the extent that the new component model impacts the model/simulation suite. As an example, assuming the new model component complies to the existing interface requirement specification, the following steps are taken for re-verification of the model/simulation suite:

- The new component is verified as described in step 1 above
- The new component and the models/simulations that it communicates with are verified as described in step 2-4 above
- Lastly, step 5 and, if necessary, step 6 are repeated for the complete simulation suite

2.3.2.3.2 Validation

(Because human performance model validation poses some unique problems, it is discussed separately in Section 2.3.2.3.2.1.)

Unlike verification, the validation phase is much more difficult to characterize in a generic sense. Each model/simulation is intended to characterize a trait of the NAS. Some of these traits lend themselves more easily to rigorous validation than others. For example, for decades now, high fidelity flight dynamics simulation of aircraft have been validated successfully against flight test experiments. On the other hand, weather and wind models have been less successful in validation when scrutinized under similar criteria. However, as mentioned earlier, central to validation is defining the scope of the issues that the model can and can not address. Designing a system that is robust to the issues/uncertainties that models can not address then becomes key.

The classic approach to validation (hence referred to as *predictive validation*) is to model a baseline scenario in which data is available (either it currently exists or can be acquired) for comparison. If the differences between the model predictions and the baseline data are unacceptable, the model is tuned/modified as needed (often referred to as calibration), re-run, and compared against the data once again. This process is repeated until the comparison between predictions and data are deemed acceptable. The model is then run for another scenario where data is available, but has not been previously utilized for comparison. If the comparison between predictions and the new data are acceptable, the model is considered validated. At this point, the validated model can be used to make predictions within a certain scope of analysis or assumptions.

The SOO adds two significant complications that challenge the process of “predictive validation”.

First, there is a requirement for multiple levels of fidelity to be available in the operation of the NAS models. However, system and individual performance data of NAS operations is recorded at specific levels of fidelity. So, in order to support validation studies, methods must be developed to aggregate performance data so that models can be validated at the level of fidelity appropriate to a particular analysis. This issue is particularly problematic in the validation of human performance, which is generally recorded at the level of individual performance. The CSC team will address this aggregation issue by developing techniques to identify the appropriate level of system performance to be used in validation studies and test data aggregation methods using historical data for the algorithms and techniques of aggregation against which to assess their validity.

The second challenge to the predictive validation technique is that the intended prediction of advanced and possibly revolutionary operational concepts, technologies and procedures may not have any appropriate data associated with their performance. In this case, the development of models and preparation will be undertaken in close connection with the NASA and contractor engineers and scientists who will be developing real-time simulation and other analysis tools in VAST. The intention is to assure that the model development under VAST NRT-M&S can be validated by the use of VAST real-time simulation technologies. In this way, the models and simulation facility development will be coordinated to be able to provide model validation data as the VAST simulation techniques are developed. In particular, human performance modeling discussed below will be undertaken so that the data from simulation is appropriate to the models' validation. This will require coordinated development and communication between the VAST NRT-M&S team and the VAST real-time simulation teams.

Since the goal of developing AMSAT is to predict NAS performance, the primary goal of AMSAT validation is to be able to state with objectivity that those predictions are accurate. The following steps describe the method for validating AMSAT:

1. All models/simulations to be considered usable components of AMSAT are validated according to predictive validation described above.
    a. For models/simulations where insufficient data, due to cost or other practical limitations, precludes predictive validation, approaches similar to those described in Section 2.3.2.3.2.1 will suffice.
2. If feasible, all AMSAT studies will begin with a baseline study that represents today's operations where data (e.g., ETMS) can be collected and compared to the AMSAT predictions. Analysis will be conducted to determine why differences exist between predictions and data.
    a. If the differences are explainable and fixable, AMSAT will be calibrated appropriately per the predictive validation approach. AMSAT, for a given scope of analysis, is then formally validated.
    b. If the differences are explainable, but not feasibly fixable, analysis will be performed to estimate how the differences can be extrapolated to modify the AMSAT predictions for studies beyond the baseline. AMSAT, for a given scope of analysis, is then formally validated.
    c. If the differences are not explainable, AMSAT will be considered not validated. This does not necessarily preclude using AMSAT for studies beyond the baseline, but NASA managers and researchers must use discretion when using AMSAT results for decisions regarding future investments into concept, tool or technology development.

AMSAT V&V will be key to every study undertaken by the AATT or VAM projects and will be documented in the Software Demonstration and Validation Report for Phases 1-4

(Phase 1 will include verification, but not validation) that corresponds to CTOD #9, 19, 30 & 41. Extrapolation of AMSAT beyond its calibrated performance is a significant issue that must be considered as specific examples arise. Studies that involve revolutionary concepts or technologies may be beyond the scope of analysis for a formally validated AMSAT. In that situation, approaches similar to the alternative validation techniques described in the next sub-section may be the only feasible strategy for AMSAT validation.

2.3.2.3.2.1 Human Performance Model Validation

Human performance model validation poses some unique problems in comparison to validation of other types of models. First, and most important, is the high degree of variability of the human as compared to other disciplines. Human performance among qualified human operators can sometimes differ by as much as 100% and typically will vary 20-40% (Reference coming from Ron Laughery later in week). Therefore, a large sample of human performance data is required to get a stable estimate of expected human performance that can be compared to a model's predictions of that performance. Additionally, human performance data tends to be difficult and expensive to collect. Collectively, this means that traditional predictive validation studies for validating human performance models will be rare. For these rare cases, predictive validation will be employed for the AMSAT human performance models. However, we fully expect that many of the human performance models will lack predictive validation for their intended application/purpose in AMSAT. As a substitute for predictive validation, other techniques will be utilized, as described below:

- *Peer validation* – Our requirements for model development include the integration of many models and modeling architectures to provide as complete and comprehensive as set of human performance modeling tools as possible. In service of that requirement and in support of model validation, the CSC team will seek review of model development from other modelers in human performance, cognition, team, and organizational performance. Specifically, NASA APEX model development will provide cognitive models of controller and pilot performance at a specific level of fidelity. The CSC team will welcome the APEX team as an appropriate judge of AMSAT human performance model development and implementation at other levels of fidelity. For internal peer review, the CSC team will leverage the expertise of three team members (MA&D, Kevin Corker-SJSU, and BBN).
- *Construct validation* – In so far as human performance is represented as a set of information processing and cognitive management steps, we propose to seek validation of the component models in empirical research in related fields of human endeavor.
- *Historical validation* – Identify data associated with the models being developed that predict performance of known and previously studied human systems. Identify the similarities between present model performance and provide estimates of the generalizability of the models to the previously studied systems.
- *Model structure validation* – For certain human operator activities, if the activities are well-understood in the form of a task analyses, then task networks models (e.g., IPME and Micro Saint) can be built to exactly replicate the task analyses.

> With proof that the underlying model principles are sound, then reasonable confidence in the model can be gained.

Human performance model validation for this CTO will formally document the validation techniques that apply to the major components of each human performance model component of AMSAT. Human performance model validation will be included as part of the Software Demonstration and Validation Report beginning with Phase 2 for the VAM option or Phase 3 for the AATT option and continue through Phase 4 in both cases. Most likely there will be significant overlap for these different techniques of validation. However, none of these techniques, collectively or individually, can ensure that a model is making accurate predictions to the extent the predictive validation can. Therefore, coordinated development and communication between the VAST NRT-M&S team and the VAST real-time simulation teams is essential to gleaning relevant validation data whenever possible.